

MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies

This policy and guidance provides detailed requirements and expectations for individual studies to meet the overarching [MRC Policy on Research Data Sharing](#) - the council's principles for data sharing that apply to all MRC-funded research. This policy and guidance was drafted specifically for the population health sciences and population and patient cohorts.

Studies with ongoing data collection or analyses should be able to show, **by 31 May 2012**, progress towards meeting the [requirements](#).

"Data are at the heart of MRC's ability to improve the understanding of human health. To maximise the exploitation of MRC data sets, it is important that MRC plays a strong leadership role in the development of informatics and infrastructure that enables effective use of MRC data."

Medical Research Council Strategic Plan 2009-14

1. Introduction

Research to understand human health and to evaluate interventions to improve health is dependent on information about the health, lifestyle, genetics and social, economic and physical environment of populations and patients. Such data provide many opportunities for collaboration across diverse research disciplines. Longitudinal studies, with repeat observations of cohorts of participants - often over many decades, are particularly rich resources for multi- and interdisciplinary investigations of development, ageing and the effects of early circumstances on health in later life.

The data involved are diverse and are collected by a variety of different means, including survey questionnaires and direct measurements on people, tests on biomedical samples, and clinical and other records.

The value of information collected from study participants grows as data are organised, 'cleaned' and quality controlled, analysed and the outputs of the analyses are made accessible to research data users. Considerable value is created through the data lifecycle.

Creating the value in these data represents a substantial commitment by the researchers, study participants and funders involved. Many players have an interest in well-managed sharing of high value research data.

The MRC and other leading research funders actively promote research collaboration and data-sharing, with the aim of maximising the value of these resources for the public good.

Researchers share rich data resources in a variety of ways. For instance, by answering new questions with existing data, validating a finding from one study by attempting to replicate the finding in another, and combining the power of individual studies that share common features through data linkage or meta-analysis.

MRC's overarching aim for data-sharing is *to maximise the life-time value of research data assets for human health and to do so timely, responsibly, with as few restrictions as possible, in a way consistent with the law, regulation and recognised good practice.*

"Research data are a valuable resource, usually requiring much time and money to be produced. Many data have a significant value beyond usage for the original research."

Managing and Sharing Data – Best Practice for Researchers. UK Data Archive, May 2011.

Purpose and scope

This policy and guidance was drafted specifically for the population health sciences and population and patient cohorts. It is based on key principles widely recognised as applicable to publicly funded research in general: [OECD Principles and Guidelines for Access to Research Data from Public Funding](#) and [RCUK Common Principles on Data Policy](#).

The [requirements](#) should also readily apply to clinical trials, while recognising that trialists may already have satisfactory arrangements in place (e.g. for trial discovery) that the *expectations* in this guidance do not seek to perturb.

The guidance has been prepared specifically for study directors, informaticians, data managers and peer reviewers; to enable the MRC research community to meet MRC policy requirements and expectations. It is the responsibility of the study director or unit director to meet the requirements for his/her studies. Units may develop a single set of measures for all their studies.

The guidance does not provide technical guidance for data managers on how to manage data or which data standards to use. Links to such information are, however, provided.

The guidance does not replace the need for investigators, data managers and others to use professional judgement and draw on other appropriate sources for advice.

Benefits of data sharing that scientists and data managers value

- ⌚ Enabling new research questions to be answered in existing data
- ⌚ Promoting collaboration between different research teams and diverse disciplines
- ⌚ Sharing of knowledge about best methods for data collection, linkage and analysis
- ⌚ Ensuring that collected data are cleaned, well documented, with value added
- ⌚ Independently verifying established research findings
- ⌚ Development and testing of new research methods
- ⌚ Using to best effect the gift of data made by study participants

Data sharing therefore represents an efficient use of public money and supports more timely scientific discovery.

Studies may share their data by archiving their data collection (or a subset) at a discipline-based repository like the UK Data Archive (www.data-archive.ac.uk), or at an institutional repository that can preserve data and make them available to users. This may be suitable for legacy data collections and studies that no longer actively collect data or receive funding.

2. Data sharing requirements for population and patient studies

The following 21 requirements are mandatory for MRC-funded cohort studies and are likely to be readily applicable to MRC population and patient-based research more broadly.

Studies should consult each topical section, which details:

- **Requirements:** mandatory principles for MRC funded population and patient cohort studies (and likely to be applicable to MRC research more broadly)
- **Expectations:** the kind of evidence MRC is looking for in assessing compliance with this policy
- **Further good practice:** recommendations for additional data sharing measures
- **Resources:** authoritative and relevant information

The [In practice](#) section illustrates how studies can put specific topical aspects into practice. Studies should use their best judgement, authoritative good practice and advice in adapting the expectations to their particular circumstances.

[Data standards](#)

R1. Studies must take an active and collaborative approach to ensuring timely development and adoption of appropriate data standards to support high quality research and data-sharing.

[Study policy on sharing](#)

R2. A simple study policy on data-sharing must be readily discoverable by the research community on the study website, in a manner sensitive to the interests of participants. The study policy must be consistent with MRC's overarching policy on data-sharing and preservation.

R3. The priorities and criteria for sharing and access, and the various constraints, must be transparent and clearly justified. The type and extent of privileged use by the study team must be clearly defined and justified in relation to the funder-approved research programme.

R4. When designing the study and seeking participants' consent and ethics approval, the Director should aim to promote the widest range of possible good uses of the data and seek to establish broad and enduring consent for data-sharing. If consent has not been obtained for data sharing, sharing can still take place if data are anonymised to the standards of the "[ICO Anonymisation: managing data protection risk code of practice](#)", and usage is broadly in line with what participants expect their data to be used for.

R5. The study must have robust policies for managing confidentiality and for data security, consistent with legal, good practice and MRC policy requirements.

R6. MRC encourages researchers to work in productive, equitable partnerships, e.g. with medical charities and industry. Sharing involving commercial or non-UK based organisations must conform to the same principles and practices as that required of the academic community.

R7. Directors/PIs must ensure that intellectual property relating to the value they create is suitably protected and managed, in line with RCUK Knowledge Exchange Principles. Any delays or restrictions on sharing due to managing IP must be minimised as far as possible.

[The governance of data access](#)

R8. Study governance of access, the criteria and processes, must be appropriate and proportionate to the nature and scale of the study, the level of risk and the likely demand for access.

- R9. The access process must include independent advice and/or oversight.
- R10. Directors shall ensure that the criteria and processes governing access are transparent and readily discoverable.

Facilitation

- R11. The principal stages and decisions in providing access and facilitating use are clearly documented, and with effective mechanisms for informal enquiry and timely feedback.

Data-sharing agreements

- R12. Directors/PIs shall ensure that a data-sharing agreement is issued and signed by appropriate authorities before data are released or analyses are performed on behalf of the requester.
- R13. Data-sharing agreements must prohibit any attempt to (a) identify study participants from the released data or otherwise breach confidentiality, (b) make unapproved contact with study participants.

Data preparation and transfer

- R14. Directors / PIs shall ensure that measures are in place to protect the confidentiality of study participants and the security of data sets when they are shared with, or analysed on behalf of, new users, and that practice complies with legal and regulatory requirements, MRC policies and relevant best practice.
- R15. Studies must ensure that metadata documentation, a metadata catalogue or personnel with relevant knowledge and expertise can support the reasonable understanding and use of study datasets by new and external researchers.
- R16. Studies must document data transfers and ensure that the data and accompanying documentation (metadata) are prepared to the agreed standards.

Funding

- R17. Funding proposals to MRC should differentiate in broad terms between the proposed costs of (i) collecting and cleaning new data and the associated cohort costs; (ii) the study team's proposed research programme; (iii) ongoing data curation and preservation; (iv) data-sharing
- R18. Unless otherwise approved by MRC, studies should not seek to generate revenue through sharing

Recognition

- R19. Studies shall promote appropriate acknowledgement of the significant contributions of all parties to creating new value through data-sharing.

Reporting

- R20. Directors / PIs must be able to report to MRC as a funder on the performance and outputs of sharing achieved during a given period of funding.

Discovery of MRC-funded studies

- R21. An MRC study that is collecting data (or has collected data) must be readily discoverable by the research community for the purposes of new research, and presented in a manner that is sensitive to the interests and continued support of the cohort participants.

3. Data standards for sharing research data

Data standards aim to promote interoperability and good data management.

Requirements

R1. Studies take an active and collaborative approach to ensuring **timely development and adoption of appropriate data standards** to support high quality research and data-sharing.

Expectations

1. Studies adopt and proactively champion a standards-based approach to managing and sharing research data.
2. Studies prioritise the compilation of well-structured variable and metadata catalogues (a) as part of good research and data management and (b) to enable data-sharing and linkage.
3. The study description and variable metadata catalogues are published to the wider research community, including where possible through the [MRC Research Data Gateway](#), to maximise the visibility to the wider research community for bona-fide research.
4. For significant datasets of data collected in the past (“legacy data sets”) that have not been fully documented, Directors/PIs plan their data management to maximise value from the data over a reasonable time period. This may involve selective documentation of variables of likely interest rather than global documentation of all variables.

Resources

The [Data Documentation Initiative](#) (DDI) is an international metadata standard to describe data from the social and behavioural sciences across the life cycle. Expressed in XML, the DDI metadata specification supports the research data life cycle, from data conceptualization, collection, processing, distribution, discovery, analysis, repurposing to archiving. The DDI community also provides [open source tools](#).

The DDI metadata standard has been used for the following population health studies:

- [Secure Epidemiology Research Platform](#) (SERPent), MRC Centre of Epidemiology for Child Health
- Norwegian [Tromsø study](#)

The [Microdata Management Toolkit](#) developed by the World Bank Data Group for the International Household Survey Network aims to promote the adoption of international standards and best practices for microdata documentation, dissemination and preservation. It includes a DDI/Dublin Core compliant Metadata Editor and Nesstar Explorer to read metadata files.

[Nesstar Publisher](#) provides data and metadata conversion and editing tools to edit and create DDI documented datasets and to prepare metadata and data for publication to a Nesstar Server.

The UK Data Archive provides researchers with [guidance on managing and sharing data](#), including guidance to documenting data and metadata. A helpful guide is their [Managing and Sharing Data: Best Practice for Researchers](#).

The Inter-university Consortium for Political and Social Research (ICPSR) provides guidance on data preparation, access, curation and archiving; e.g. through their [Guide to Social Science Data Preparation and Archiving](#).

The [Australian National Data Service](#) provides a valuable guide to data management for researchers, including an introduction to metadata (data about data).

The [National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation](#) sets out how producers of National Statistics will carry out their responsibilities for managing, documenting, retaining and preserving the statistical resources they control.

4. Study policy on sharing research data

Individual studies articulate their own specific data-sharing policies, in terms of the purpose of the study, their sharing priorities and their partnerships.

Requirements

R2. A summary of the principles of the **study policy on data-sharing** must be readily discoverable by the research community on the study website, in a manner sensitive to the interests of participants. The study policy is consistent with [MRC's overarching policy on data-sharing and preservation](#).

R3. The [priorities and criteria for sharing and access](#), and the various **constraints**, are transparent and justified. The type and extent of [privileged use](#) by the Principal Investigator and study team are clearly defined and justified in relation to the funder-approved research programme. Access to study data is on a *non*-exclusive basis.

R4. When designing the study and seeking participants' consent and ethics approval, the Director aims to promote the **widest range of possible good uses** of the data and seek to establish **broad and enduring consent** for data-sharing.

R5. The study has robust policies for [managing confidentiality and for data security](#), consistent with legal, good practice and MRC policy requirements.

R6. MRC encourages researchers to work in productive, equitable **partnerships**. Sharing involving commercial or non-UK based organisations conforms to the same principles and practices as that required of the academic community.

R7. Directors/PIs ensure that **intellectual property** relating to the value they create is suitably protected and managed, in line with [RCUK Knowledge Exchange Principles](#). Any delays or restrictions on sharing due to managing IP are minimised as far as possible.

Expectations

Transparency

1. The summary (or whole) policy, or the link to a downloadable copy, is easy to find, and no more than one mouse-click from the home page on the study's website.

Study purpose and data-sharing priorities

2. The policy helps potential new users understand the **value of the data** and the kinds of new use for which sharing is **likely to add value** and any **constraints** on when or purposes for which access can be provided.
3. Where possible the **acceptability** of the key principles of the study policy on data-sharing to the study **participants** has been ascertained.
4. The principles of the policy are considered and approved as part of the data-management plan at **peer review**.
5. The policy provides information on:
 - the [purpose](#) of the study
 - specific [opportunities and priorities for sharing](#)
 - the terms of any [privileged use of data](#) by the study team
 - the standard use conditions for [secure management of confidential and sensitive information](#)
 - how the [costs of sharing](#) are met.

Data lifecycle and quality considerations

6. Potential new users can identify the period occupied for data preparation (e.g. collection, 'cleaning' and quality control by the study team) before data are ready to be analysed and before they are ready to be shared.
7. A good policy recognises where the strengths and weaknesses of the study data lie, and where additional work and funding would be required to bring data up to the quality necessary for sharing. Experienced researchers understand that data can be "messy" and that data collected in the past ("legacy data") may not be organised optimally, fully 'cleaned' or adequately described with metadata. Some legacy data may have limited value.

Consent and other ethical considerations

8. The sharing policy indicates the ethical principles and conditions that promote (or constrain) the nature and extent of data-sharing, for instance:
 - the terms of existing **participants' consent** and ethical approvals pertinent to sharing of data already collected
 - whether broad and enduring consent will be sought for **future sweeps** (collection of new data)
 - sharing of data **outside the UK**, including outside the EU
 - whether related **genetic data** and **biomedical materials** may be made available.

For established studies, the extent to which ethics committees will approve broad consent for future new uses may be constrained by the consent(s) given by participants in the past, which may be rather vague or restrictive (e.g. "...not shared outside the research team"). For new sweeps of existing studies, and new studies, it should be reasonably straightforward to establish consent for sharing such that participants' data are used to maximal good effect.

Where data sharing has not been mentioned in consent documentation or discussions, the absence of consent is not a reason not to share; as long as data can be anonymised in line with ICO standards and usage is broadly in line with participant expectations, explicit consent for sharing is not required.

Capability and capacity to use the data

9. A study may give priority to new uses that are for [bona fide](#) research.

Partnerships

10. The policy states the principles of partnership and how partnerships are managed; this may be partnerships with individuals, organisations and industry. Critical obligations are documented in data-sharing or other appropriate [agreements](#).

The costs of sharing

11. Studies set out their policy for meeting the [costs of sharing data](#).

Further information

12. The policy indicates what **further information** is available either on request or through a hyperlink, and indicates a [contact point](#).

Further good practice

Institutional policies

Institutions hosting MRC studies have robust and compatible policies for handling confidential data and for information security.

- f* A confidentiality policy that specifies how the identities of, and knowledge about, study participants are protected from unintended disclosure
- f* Policies that specify how data are handled and protected within a study, addressing information security and statutory and regulatory requirements (e.g. Data Protection Act 1998, health department Research Governance Frameworks).

Resources

Examples of MRC-funded studies with online data-sharing policies include:

- f* The [Avon Longitudinal Studies of Parents & Children](#) (ALSPAC)
- f* [Whitehall II](#), a study of social class, psychosocial factors and lifestyle as determinants of health and ill health.

The UK Data Archive (UKDA) who manage the Economic and Social Sciences Data Service, provides researchers with guidance on managing, preparing, storing and sharing data. Especially useful may be their guidance on:

- f* [Ethics, consent and data sharing](#)
- f* [Data security](#)
- f* [Quality assurance](#)
- f* [Copyright](#) of research data

Data Sharing for Statistical Purposes, a practitioner's guide to the legal framework, Office for National Statistics (2005), includes a code of practice for processing ONS data and a useful explanation of legal terms and concepts.

[UK Biobank](#), a study collecting medical and genetic data from 500,000 middle-aged people across the UK to create an information resource to study the prevention and treatment of serious diseases, has detailed protocols for ethics, governance, access and intellectual property.

The [Intellectual Asset Management for Universities](#) guide of the Intellectual Property Office gives advice and information to universities to help them understand how they can best use their institution's intellectual property.

[Copyright and Intellectual Property Law](#) by JISC Legal Information Service provides guidance, examples and FAQs on intellectual property rights such as copyright, patents, trade marks, design rights, and the protection of confidential information.

[IPR and Licensing issues in Derived Data](#) by Naomi Korn, Professor Charles Oppenheim and Charles Duncan analyses the issues of IPR and licensing in text and data mining.

The University of Cambridge provides a useful [guide to data and IPR](#).

The Digital Curation Centre [Legal Watch paper on Sharing Medical Data](#) explores the legal consideration for sharing data that contain personal information.

The Research Councils UK [Knowledge Exchange Principles](#) explain their position on intellectual property and asset management. The responsibilities of Research Council grant-funded organisations are set out in the [Research Council Terms and Conditions](#) (Paragraph GC 21 - Exploitation and Impact).

5. The governance of data access

This section of the Policy and Guidance on Sharing of Research Data from Population & Patient Studies covers the criteria and processes for governing sharing and access requests for research data.

The [graph below](#) illustrates the possible principal steps and work flows in processing a data-sharing or access request, based on existing MRC cohort studies, without prescribing a standard workflow.

Requirements

R8. Study governance of access, the criteria and processes, are **appropriate and proportionate** to the nature and scale of the study, the level of risk and the likely demand for access.

R9. The access process includes **independent** advice and/or oversight.

R10. The **criteria and processes governing access** are transparent and readily discoverable.

Expectations

Transparency

1. How a study manages **access** is transparent, informative and readily discoverable, in a manner sensitive to the interests of the cohort participants.

- f* Summary information on the study website covers the [criteria by which sharing/access requests are assessed](#) and the processes and timeframe by which the requests are assessed.
- f* Additional details on how data selection and use is [facilitated](#) by the study team are readily available on the website.

2. Governance is **effective, appropriate and proportionate** to the following:

- f* Anticipated demand, e.g. number, scale and complexity of data-sharing requests
- f* The lifecycle stage of the study (collection, PI-led primary analysis, curation of the data as an available resource)

- f* Risks related to the sensitivity of the data (e.g. the potential for harm and distress to participants from the release of certain data or deductive disclosure)
- f* The time period over which the data are expected to have research value (especially for genetic analyses data can become outdated as new assay techniques are developed)
- f* The anticipated scale of study team facilitation and resources required to meet the request.

Roles and responsibilities

3. The Study policy defines the [role of the Director/PI and study staff](#) in decisions about access and facilitating sharing. It also sets out the responsibilities on external users (which are reflected also in the Data-sharing Agreement).

Independent advice and oversight

4. The access governance process is subject to independent oversight of (a) external requests for data (and related materials, such human tissue and DNA, if appropriate); and (b) new investigation requests from within the study team for uses that are not already approved by MRC (or other relevant funder or sponsor) as part of the study's research programme.

5. The terms of reference and working method of the reviewers are transparent. [Two possible models](#) for the review process are given below. The rules of public accountability operate, such that an external advisor plays no part in decisions in which they have an interest by virtue of their professional or other relationships.

6. The review process may provide for additional, external advice on specialist topics if the need arises.

7. The policy sets out the study arrangements for re-review or other form of internal [appeal](#) after a decision not to provide the data (or not as requested). The policy includes provision for the details of a request, the review process and outcome to be referred to the study's MRC programme officer in the event that a review of the decision and process by the independent committee or advisor does not resolve concerns or dissatisfaction.

8. To achieve efficiency and consistency, studies may choose to share a common oversight committee (or advisor). In doing so they should consult their MRC programme manager.

Resources

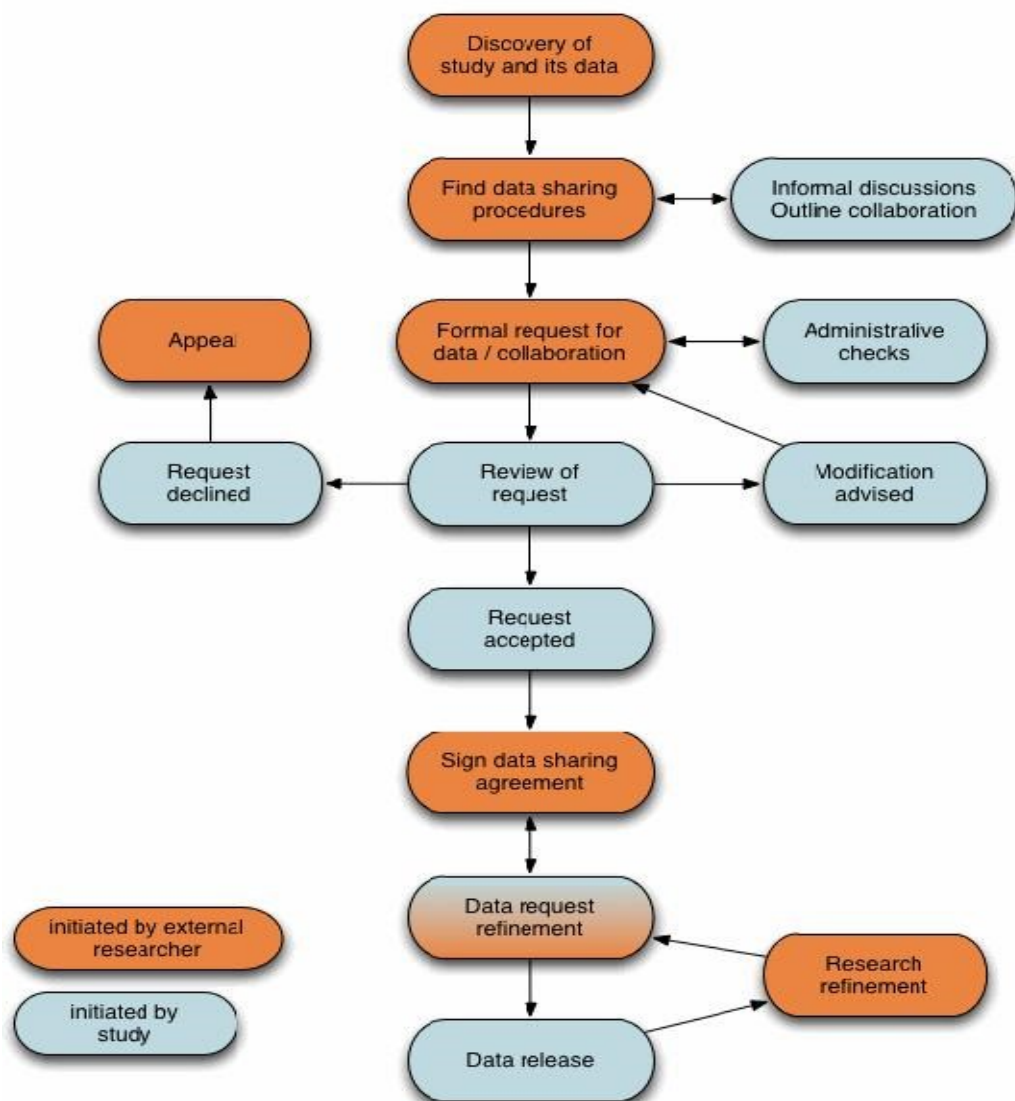
The eight [principles of the Data Protection Act](#) (1998).

The NCRI has created a [template for access policy development](#) for data and tissue samples.

The Digital Curation Centre has developed guidance on how to [Appraise and Select Research Data](#) for curation and sharing.

[Guidance on controlling access](#) to shared research data is available from the UK Data Archive.

Data-sharing request workflow



6. Facilitating data-sharing

Requirements

R11. The principal stages and decisions in providing access and facilitating use are clearly documented, and the mechanisms for informal enquiry and timely feedback are effective.

Expectations

Point of contact

1. There is a contact mechanism for questions about the study, data sets and data-sharing through e-mail.

- f* Enquiries via the contact point are documented, so that studies can ensure that enquirers are satisfied that their enquiry has been fairly handled, and can monitor the efficiency of the access process.

Practical information available on the study website

2. The study provides **information** or links to the following:

- f* A clear visual guide to the [data-sharing process](#)
- f* A data-sharing request form, structured according to the criteria for access
- f* An indication of user requirements, as are specified in the study's standard data-sharing agreement(s).

3. **Practical "how to" advice** is available, for instance on:

- f* How to **select** study variables of potential interest from the [Population and Patient Research Data Directory](#) or from accessible documentation of study variables, with at least guidance on those variables most likely to be of value for new uses
- f* **Consent and other ethical requirements** needed for new uses, and how and when permissions should be sought
- f* **Data security** requirements of external users
- f* How data requests should be [costed](#).

Informal discussion prior to submitting a funding request

4. A study may recommend that external researchers consult the study before submitting a research funding proposal that would involve use of the study dataset.

- f* The purpose is to verify that in the event of an award, the data are likely to be available, and the request is otherwise likely to meet the study policies on sharing and access.
- f* The process could mirror internal requirements for study team members to receive their director's approval to develop significant new projects requiring new uses of the data.

Formal request for data access

5. The study provides a formal request form (and/or checklist) designed to elicit the information required for review, in a format compatible with the governance criteria and workflow.

6. On receipt by the study, the request is documented. The request form (and accompanying documentation) is checked to ensure it is appropriate and complete, before initiating the formal review. Receipt is acknowledged, with feedback on whether the request is suitable for formal review.

Informative feedback about a decision

7. Summary feedback is provided within a few days of the formal decision and, if necessary, is followed up with concise written details.

8. For declined requests, the grounds for rejection are explained in relation to the access criteria. The feedback explains what modifications to the research protocol might enable the request to be approved. Options for re-submission are specified.

9. For approved requests, the feedback specifies and explains the following:

- f* Whether there are conditions or limitations to the kind of data that can be made available, for example if certain variables cannot be provided
- f* Whether certain parts of the data set will need cleaning prior to use, and whether the requester is expected to assist in this in person or through financial resources
- f* Precise requirements in relation to existing ethics approval; whether further approval(s) will be required and the roles of the study and data requester in applying for further approval(s)
- f* Funding implications for data use and collaboration and any costs to be charged for data preparation
- f* The next steps, including any time constraints, such as a time limit for access to, or use of, the data
- f* Arrangements for access to samples or the process for adding questions to a future questionnaire, if applicable
- f* Conditions that will be specified in a [data-sharing agreement](#).

Further good practice

1. A response time is indicated for informal and formal requests. Responses are timely and informative.
2. Studies make appropriate provision for staff to interact with requesters to explain variables in detail, including their context of collection, closely related variables, why particular data quality was observed, the evolution of a variable between collection waves.
3. Each agreed data-sharing relationship has a nominated contact point within the study team whose responsibility is to monitor the progress and outcome of the external research.
4. An escalation procedure enables that contact point to alert senior study personnel if problems or delays arise in the external research or if there are concerns about compliance with the data-sharing agreement.

7. Data-sharing agreements

Requirements

R12. Directors/PIs shall ensure that a data-sharing agreement is issued and signed by appropriate authorities before data are released or analyses are performed on behalf of the requester.

R13. Data-sharing agreements must prohibit any attempt to (a) identify study participants from the released data or otherwise breach confidentiality, (b) make unapproved contact with study participants.

Expectations

1. The data-sharing agreement specifies the following:
 - f* The parties to whom data are released. Both the (bona-fide) researchers and their institutions should be specified with precision.

- f* The specification of the dataset(s) to be prepared and released, and/or the analyses to be run securely on behalf of the requester.
 - f* The purposes for which data are released or analysed (annexing a copy of the research protocol for the new use).
 - f* The conditions under which the data may be used, particularly in relation to ethics committee approvals.
 - f* The specific obligations and arrangements to maintain confidentiality and data security.
 - f* The handling of intellectual property, publication, authorship, acknowledgement and whether data are provided on an “exclusive” or “non-exclusive” basis to the requester.
 - f* Any constraints on publication related to the study’s privileged use of its data or to manage specified risks.
 - f* A requirement that research publications and other outputs based on the transferred data (or analyses conducted by the study on the new users' behalf) are reported to the study.
 - f* A requirement for the study and MRC to be appropriately acknowledged in publications and other outputs.
 - f* Other requirements in relation to preventing or controlling onward transfer of the data (or the data derived thereof by the requester) to a third-party, and for conferring to third parties the appropriate obligations for custodianship, use and reporting).
 - f* A requirement to notify the study systematically of all research publications based on the use of the study data, and similarly other information needed to allow the study to report on its data-sharing performance and outputs (see Section 11).
 - f* Whether a copy of data derived by the requester must be made available to the study, possibly after a reasonable delay (i.e. a period of exclusive use by the external user). It may be stipulated that the derived data are to be retained by the external user.
 - f* Individual roles and responsibilities in relation to supporting the agreement, including arrangements for ongoing support from the study to facilitate use of the data.
 - f* Arrangements for data destruction or secure archiving.
 - f* Any costs that need to be contributed by the requester
2. The agreement reflects the nature of the relationship between the study and the new user, which can range from direct provision of data, data analysis collaboration and/or scientific collaboration.
 3. Roles and responsibilities are appropriately allocated. IPR, [recognition](#) (authorship and acknowledgement) and other benefits are distributed according to the specific contributions of the individuals and organisations involved.
 4. The agreement is signed by authorised organisational representatives from both parties. Organisations who have signed the data-sharing agreement ensure that users of shared study data fulfil the obligations stipulated in it.

Further good practice

Data-sharing Agreements may also specify the following:

1. A time period for which the approval has been granted
2. Whether the external researchers should report progress.
3. How the originally supplied datasets, subsequently derived data and analysis files should be managed, in accordance with ethics committee stipulations and in proportion to the confidentiality risks.
4. An obligation on data recipients to commit to and apply security and confidentiality measures to the study data they receive (and derive from the data) that are equivalent to those under which the source dataset are held; and informing a study of any incidents of breach or unauthorised disclosure.
5. The research endpoint that should trigger secure archiving and/or destruction of data.
6. A requirement for the requester to provide the study with evidence of their security and of data destruction policies.
7. An obligation to delete the records of any participant who has withdrawn from a study if they are identifiable.
8. Whether pre-publication permission is required from the study and the scope of any constraints on publication, e.g. a limited publication moratorium to allow the study to publish the results of the funder approved programme. Agreements should specify a rapid-response mechanism for pre-publication approvals.
9. The criteria on which study team members, including study data managers, informaticians and statisticians, should be included as co-authors, or otherwise recognised, in publications and other outputs derived from the study data.
10. A requirement that publication is consistent with the [RCUK position statement on Open Access](#).
11. Agreements should include any constraints and licensing requirements specified by the funder or organisation with custodianship of the data.

Different data-sharing agreements may be needed for different types of data, e.g. for anonymised versus disclosive data.

Resources

Example data-sharing agreements are available via the Regional Contracts Manager at MRC Regional Centres.

[Open Data Commons](#) has a set of tools to enable provision and use of 'open data,' suitable perhaps for summary data tables of 'unrestricted availability.'

The NCRI has created templates for access policy development, which includes a [template Data and Material Transfer Agreement](#)

The Digital Curation Centre has guidance on [How to License Research Data](#).

8. Data preparation and transfer

Requirements

R14. Directors / PIs shall ensure that measures are in place to protect the confidentiality of study participants and the security of data sets when they are shared with, or analysed on behalf of, new users, and that practice complies with legal and regulatory requirements, MRC policies and relevant best practice.

R15. Studies must ensure that metadata documentation, a metadata catalogue or personnel with relevant knowledge and expertise can support the reasonable understanding and use of study datasets by new and external researchers.

R16. Studies must document data transfers and ensure that the data and accompanying documentation (metadata) are prepared to the agreed standards.

Expectations

1. Studies create and retain a record of how they prepared the data, which data have been transferred, when, via what media, and whether data were encrypted.
2. Studies document significant resources they devote to data preparation, metadata, other kinds of documentation and transfer.
3. Studies should aim to provide requested data in accessible or standardly used file formats.

Further good practice

1. The extracted data set is checked for small cell sizes (counts) and appropriate action taken (for example, values might be merged or blurred or suppressed).
2. A clear policy specifies when pseudo-identifiers should be used and when study/sample identifiers can be re-used.
3. Where this enhances data security, studies replace study/sample identifiers with unique pseudo-identifiers to limit the risk of re-identification.
4. These pseudo-identifiers are normally re-used for supplementary releases to the same external party so that recipients can link successive released data sets and to facilitate repeat access to samples in the future.
5. Specific data releases can be identically regenerated if necessary, to enable reproduction of particular results or for research governance investigations.

Resources

The eight [Principles of the Data Protection Act](#).

[Secure safe haven architectures](#) can facilitate containment of datasets within a controlled and secure environment, reducing unnecessary replication and enabling detailed audit of data access.

Tools that support deductive disclosure risk assessment and mitigation include:

- [Anonymisation - Measuring the disclosure risk](#) of the International Household Survey Network.
- [SUDA - a program for detecting special uniques](#) by the University of Manchester.

The **UK Data Archive** has guidance on preparing research data for sharing, in particular:

- f* [Anonymising data](#)
- f* [Documenting data](#)
- f* [File formats](#)

9. Funding data-sharing

Peer reviewers of study proposals frequently find difficulty in differentiating between the costs of the study team's research and the costs associated with maintaining the cohort and the data. Without transparency, the costs of the research can look very expensive. Better differentiation between the cost drivers should enable funding applicants to better justify their proposed costs.

Generally, the cost of building in good practice prospectively is significantly less than when trying retrospectively to address the curational and sharing needs. New studies can be expected to have planned effective, efficient and economical means for managing data, including the processes supporting sharing.

Not all legacy datasets necessarily merit retrospective investment to make them available for new use. The quality and relevance of the data, the likely demand, and the costs of bringing them up to a shareable standard, are all factors that should be taken into account.

Requirements

R17. Funding proposals to MRC should differentiate in broad terms between the proposed costs of (i) collecting and cleaning new data and the associated cohort costs; (ii) the study team's proposed research programme; (iii) ongoing data curation and preservation; and (iv) data-sharing.

R18. Unless otherwise approved by MRC, studies should not seek to generate revenue through sharing, but may recover the [costs to prepare data for sharing](#).

Resources

The [Keeping Research Data Safe](#) (KRDS) website contains advice on cost / benefit analysis, particular in relation to research data preservation to enable ongoing access in the UK higher education sector.

10. Recognition

All new users of shared data must appropriately acknowledge the original study in any reported or published findings. The reward system needs to include all those who create value – the scientists, analysts, informaticians, data managers and others.

Requirements

R19. Studies promote appropriate acknowledgement of the significant contributions of all parties to creating new value through data-sharing.

Expectations

1. [Data-sharing agreements](#) specify the criteria for co-authorship, consistent with established community practice.
2. Study informaticians, statisticians and data managers and others who play a key role in enabling data-sharing and facilitating successful use of shared datasets, including through the development of standards and tools taken up by others, are recognised appropriately.

Further good practice

Digital Object Identifiers (DOIs) to reference datasets uniquely might prove a useful method of finding study citations within publications. Information on DOIs is available from [DataCite](#).

Resources

[Vitae Researcher Development Framework](#)

The Digital Curation Centre have guidance on [Data Citation and Linking](#) [RCUK information on Open Access](#)

[MRC Good Research Practice](#), the principles of good research practice that all MRC-funded scientists are required to follow as a condition of their funding, provides guidance on data storage and back-up, the retention of data, the reporting of findings and the acknowledgement of contributors.

11. Reporting on data-sharing

Good reporting not only meets the requirements of researchers to be accountable to the funder for how they use public funds, but also enables their institutions and the funders to celebrate the success of studies and those who make secondary use of the data.

The reporting framework aims to achieve informative, brief reporting - and comparability, through highlights of achievements and a few key metrics.

Requirements

R20. Directors / PIs must be able to report to MRC as a funder on the performance and outputs of sharing achieved during a given period of funding.

Expectations

1. The study produces a short report for its funders periodically (e.g. as part of new funding requests or quinquennial review), which includes the following information, as appropriate to the study, its stage in the study lifecycle, and the nature and scale of its data-sharing:
 - f A brief summary of the kind of **scientific opportunities** that the new uses are enabling. An indication of the kind of relationships and partnerships involved.
 - f **Process Metrics:** Number of informal requests in the reporting period; the number of formal requests; the number (%) of formal requests that were

accepted; the number (%) of accepted requests that were formally referred back to the requester for revision or substantive additional information; the number (%) that were declined; and the number (%) that appealed.

- f* **Resourcing metrics** for the funder: The cost to the study of (a) the access governance process, and (b) facilitative selection, preparation and analysis of data for the requester. The costs contributed to (b) by the requesters.
- f* **Output / outcome Metrics:** Standard research publication and other output information, indicating which were led by the study team and which by external users.
- f* Studies may provide other indicators to illustrate the **value** they are creating through sharing.
- f* Any data breaches resulting from data-sharing

2. For studies with for example eight or more unrelated requests a year, a report is regularly produced for the oversight committee (annually) along the lines above.

3. Summary highlights and metrics are also made available to participants and the wider research community, and ideally are accessible through the study website.

12. Discovery of MRC-funded studies and data

For the purpose of this guidance, discovery refers to the visibility of studies and therefore the ability for new and potential users to find out about existing studies and their variables; and to be able to judge from the information available whether the study data are suitable for their planned research.

Requirements

R21. An MRC study that is actively collecting or analysing data (or otherwise is recognised as having value for new users) is **readily discoverable** by the research community.

Expectations

1. **The study has a website that provides up-to-date summary information** about the study purpose, the cohort profile and the kinds of information (being) collected, in sufficient detail to inform the general public about the study, and consistent with the expectations and sensitivities of the study participants.

2. Once a study is collecting data, the study website contains **a technical description of the study and variables** for use by the research community, or provides a direct working link to that information.

- Published information includes a reference cohort profile and reference publications (describing the cohort and study).
- A full variable list is browsable, searchable or downloadable by *bona fide* researchers
- The study description and metadata about usable variables is made available through the [MRC Research Data Gateway](#) (once operational) and, optionally, through other means.

3. The study website and links provide the public and the research communities with **informative examples of significant research findings**. They highlight the role in

those studies and findings of cohort participants, the wider public, patient and service user group involvement, research collaboration and other forms of data-sharing and how participant confidentiality is secured.

Further good practice

1. A single main reference publication is specified for citation of the study (e.g. cohort profile in International Journal of Epidemiology).
2. Progress with any current wave(s) of collection, and expected availability of that data, is regularly updated.
3. Sample (blank) questionnaires or interview guides used for each completed wave are downloadable by *bona fide* researchers, optionally including any restrictions on copyright and re-use.
5. The variable list includes metadata such as value lists, completion rates, data quality indication etc.
6. Studies implement MRC software for “my favourites” or “shopping baskets” of commonly requested variables that can be defined and easily requested by researchers.

Resources

A prototype of the [MRC Research Data Gateway](#), implemented in 2011, supports the documentation and discovery of comprehensive variable lists and their metadata. This gateway contains a Population and Patient Research Data directory. Tools for incorporating study-specific metadata in various commonly used formats are also being developed.

Examples of MRC-funded studies with informative websites include the following:

- f The [Avon Longitudinal Studies of Parents & Children](#) (ALSPAC)
- f [Whitehall II](#), a study of social class, psychosocial factors and lifestyle as determinants of health and ill health

In both cases, the study data-sharing policy, data dictionary and questionnaires are readily accessible one-click and PDF file away from the study’s home page.

The [National Data Archive](#) (NADA) is an open source web application developed by the International Household Survey Network, for the cataloguing and dissemination of microdata documented in compliance with [Data Documentation Initiative](#) (DDI) and [Dublin Core](#) metadata standards. It can be used to build a data catalogue and portal for browsing, searching, applying for access, and downloading data and metadata; based on DDI.

[Nesstar](#) allows the publishing and dissemination of metadata and data.

13. Roles and responsibilities in data-sharing

Study directors or principal investigator, data requesters and the MRC each play a role in enabling the sharing of research data.

Study Director / Principal Investigator

Director in this guidance refers to the person with overall responsibility for an entire study, which may comprise several research programmes or projects, each with their own lead scientist. **Principal Investigator** refers to the person responsible for a research grant funded programme or project involving data collection and analysis.

The roles of a Director or Principal Investigator includes the following:

- As a **scientist and principal investigator** leading an MRC-approved research programme, they have overall responsibility to their funder(s) and employer for delivering and responsibly communicating **high quality research outcomes**, and for good research conduct, in accordance with MRC Terms and Conditions
- If the **custodian of the cohort**, they are responsible for **sustaining** the goodwill and participation of cohort participants and, in that regard, for safeguarding the reputation of the study and ensuring confidentiality in accordance with ethical and legal requirements
- As **custodian of the research data**, they are responsible for making **best use of the participants' data**, including through sharing, and for the integrity, security and quality of research data management, in accordance with institutional policies and recognised data standards
- As a **senior manager**, they are responsible for sound policies, systems and processes for deploying and managing **study resources and funding** – balancing the requirements of the research, the cohort and the data
- As a **leader**, they are responsible for **innovation and improvement**, both in their research and in the management and sharing of valuable research resources.

Data Requester

The **data requester** and **new user** responsibilities include:

- As a **scientist**, to make best use of the data entrusted to them, delivering and responsibly communicating **high quality research outcomes**, according to good research practice and the policies of their own institution and funder
- To respect the **interests of the cohort participants**, recognising their gift of personal data, the dependency of the study now and for the future on the goodwill of the participants, and respecting their wishes for use of the data, e.g. regarding confidentiality
- To ensure the **integrity, security and quality of information** entrusted to them, and of data they then derive, equivalent to those under which the source data are held and in accordance with MRC policies and recognised data standards
- To respect the data-sharing agreement(s) they have entered into as a secondary **user** and/or collaborator
- To properly **acknowledge** the original of the data and the significant **contribution** of various parties towards their creation.

In terms of research conduct, the responsibilities of the requester and secondary user are, in principle, the same as those of the study director.

MRC

The MRC's responsibilities include:

- Promoting high quality collaborative research and associated data-sharing
- Setting, promoting and reviewing policy and guidance, working with other organisations to achieve harmonisation
- Promoting the development, validation and implementation of appropriate data standards
- Auditing data-sharing study policies and practices by MRC-funded studies
- Enabling the visibility of rich data resources (through the Data Support Service project)
- Promoting professional development of MRC data managers and enabling sharing of best practice
- Compiling and reviewing evaluative data.

14. In practice

Illustrations of topical aspects to help studies meet the requirements and expectations in practice. Various sections of the Policy on Sharing of Research Data from Population and Patient Studies refer to these practical illustrations.

Study purpose and priorities for sharing

A study policy and the supporting information should include:

Study purpose: The research themes and questions for which the study is (and has been) funded, so that potential users can identify the unique potential of the study for productive new-uses.

Specific opportunities and priorities for data-sharing: Study policy in relation to maximising value through responsible sharing is clear, and specific priorities for sharing are transparent, e.g.:

- Whether priority is given for new uses that complement the study objectives and/or for which the study is a unique resource.
- Whether the study is open to unforeseen opportunities – especially for interdisciplinary synergy, research translation and knowledge transfer.
- Whether there are opportunities for non-research uses such as teaching and dissemination.
- Whether, when and how biomedical materials (e.g. human tissue, DNA) and other resources may be available for sharing.
- How requests from projects with closely similar research questions are handled; for instance, whether such groups are encouraged to collaborate with each other, and if not how competing projects are prioritised (see [access governance](#)).
- Illustrative examples of previous/current data-sharing collaborations and their most significant outputs (datasets and publications).
- Whether and how additional questions or measures can be added to future waves of data collection - including indications of the likely costs involved.

Privileged use of data by study team

Over the data lifecycle, a balance needs to be struck between the needs of the study team and those of potential new users, reflecting their actual and potential contribution to maximising the value of the data.

The intellectual and managerial efforts that go into study design, achieving funding, data collection, management and analysis need to be incentivised and rewarded. It is reasonable for principal investigators (PIs) to have specific arrangements for them to analyse the data and to publish results. These arrangements may comprise a limited period of exclusive use.

In practice, facilitated collaboration with external researchers can significantly increase the value and quality of the data even early on, e.g. through contributing to data cleaning and creating derived variables.

The study policy on sharing should define the terms of privileged use by the PIs, such that they are transparent and can be verified. If the policy provides for a period of exclusive use, it should take into account the following considerations:

- Where a study's principal investigator(s) are funded to deliver a specific programme of *research*, a *limited* period of privileged use of the data they have acquired and to which they have added value is reasonable.
- For practical reasons this time period may be indicative and need to be revised if delays occur (e.g. in recruitment). Different periods may be applied to different datasets, e.g. to take account complexity of cleaning and documentation.
- Timings will depend on a study's data collection patterns.
- However, extended privileged use by PIs – e.g. beyond that approved through peer review – could be controversial and open to challenge.
- In relation to timing, the terms could, for instance, be expressed as follows: "6-months after the end of the current grant period", "12-months after new data collection to allow for data cleaning and documentation" or "3-months following the first publication of findings based on the data."

The principles of a study's privileged use of data should be explicit in the study's funding proposals, as part of the Data Management Plan (if such a plan is required by the funder).

Categories of data availability

Studies may find the following categorisation of data availability helpful:

1. **Unrestricted availability:** Anonymised data (e.g. summary tables) for which the risk of disclosure (identification of individual participants) directly or through association with other data sources is extremely low, which can safely be made readily accessible without restriction ("public").
2. **Independently available:** Data available in principle for use by independent new, [bona fide research](#), within the terms of participant consent and not restricted by IPR, prior collaborations or other reasons, and for which the necessary metadata are well documented and available.
3. **Dependently available:** Data restricted by the scope of historically obtained participants' consent, or other reasons such that sharing (and possibly analysis)

can be effected only with significant facilitation by the study team (e.g. where data and metadata quality is not sufficient for independent use).

4. **Unavailable:** Data not available for sharing because of ethical, IPR, prior exclusive agreements or other constraints.

Bona fide research

“Bona fide” is frequently used in relation to sharing of research data, but is rarely defined.

For the purposes of this guidance, key characteristics of **bona fide research** can be considered to be as follows:

- **An intention to generate new knowledge and understanding** using rigorous scientific methods. (This includes discovery research, development and validation of methodology and technology, validating and challenging previous findings, and pilot research). *And...*
- **An intention to publish the research findings and share the derived data** in the scientific community, without restrictions and with minimal delay, for wider scientific and eventual public benefit. (Recognised constraints include a short prepublication delay to ensure proper management of intellectual property). *And...*
- The intended activities are not inconsistent with **legal and ethical requirements** or widely recognised good research practice.

In practical terms, a research project or proposal that has been approved by a recognised funder should normally be considered to be “bona fide”.

A **bona fide research organisation** is one that has the capability to lead or participate in high quality, ethical research. It will have a public commitment to adhere to recognised research and information governance good practice. (It is not a requirement that such research is the primary business of that organisation, or that all of the research undertaken by that organisation is published. Nor is it a requirement that the organisation be publicly funded.)

A **bona fide researcher** is a person with

- the professional expertise and experience to conduct bona fide research
- and
- a formal relationship with a bona fide research organisation that requires compliance with appropriate research governance and management systems.

Confidentiality and data security

Maintaining confidentiality and data security are crucial to public confidence in research.

- The study policy on data-sharing should set out the standard use conditions for secure management of confidential and sensitive information.
- The conditions should distinguish appropriately between data that are non-disclosive, potentially identifiable (although anonymised) and identifiable (not anonymised); this may relate to their [categories of data](#).
- The standards required of external users and study team members should be the same, and should be appropriate to the risks.

Study policy and practice must conform to

- [MRC guidance on Personal Information in Medical Research](#)
- [MRC guidance on human tissues](#)
- [MRC information security policy \(for MRC units and institutions\)](#)

Meeting the costs of data-sharing

The study should set out a policy for meeting the costs of sharing. Policies may cover the following:

1. Whether publishing (or having readily available) **anonymised data tables** is an effective means for supporting bona fide new use at low cost to the study and 'free' to the research community.
2. Whether the scale and costs of **data-sharing** merits building an explicit **budget-line** into the study's MRC next [funding request](#)
3. Whether **new users are asked to contribute** (e.g. through grant support) to the costs of preparing data and facilitating use (or analysing data on behalf of new users), and if so what proportion. The greater the resource needed to support new use, the more reasonable it will be to recover some or all of the costs of data preparation and facilitation.
4. Whether the study will **absorb some of the sharing costs** for proposed new use that closely complements the study team's MRC-approved research programme.
5. Whether for **commercial use** the study charges a premium over and above the costs of preparing the data for sharing, reflecting the value of the intellectual asset. Such arrangements must be on a non-exclusive basis. They should be consistent with [RCUK guidance on management of intellectual assets](#).

Commercial use of data

MRC encourages researchers to work in partnerships, also with commercial organisations.

- Studies may charge for cost recovery for data-sharing
- Studies may charge a premium for their Intellectual Property Rights (IPR)
- f* Proposed use by commercial organisations should be for [bona-fide research](#) and educational uses only, that are consistent with the public interest
- f* Exclusive relationships for the use of data should be avoided, both for commercial and non-commercial purposes.
- f* The MRC technology transfer office can advice on the charging for IPR and should be consulted by PIs.

Criteria for considering data access requests

The **criteria** for approving data access need to be transparent and appropriate to the study, the funder's intentions and the participants' interests. New use should be consistent with maximising the value of the data, across the lifetime of the study, for public good. The following should be taken into account in developing study-specific criteria:

Study purpose and priorities for sharing

- The request is consistent with the [study policy on sharing](#).

- New use is compatible with the requirements of MRC for the study team to deliver on their MRC-approved research programme.
- Where a proposed new use is similar to an existing use, the new use should be compatible with maximising the value derivable from the data
- There is not an obviously more appropriate source of data for the request.

Capability and capacity to use the data for high quality research

- The proposed research is [bona fide](#) and is methodologically sound: e.g. the analysis plan is robust and the requested sample structure and size will be sufficient to generate meaningful results.
- For research use (rather than educational use), the requestor is a *bona fide* researcher working within a *bona fide* research organisation.
- Any other new uses of subsets of the data, e.g. for teaching, are specified and similarly the requestor can demonstrate the capability and capacity to use the data.

Lifecycle

- f* Relevant data of the necessary quality are held by the study and can reasonably be made available for use (e.g. collection, validation and cleaning have been completed).
- f* Sharing of the data now does not obviously compromise the ability to derive additional value later.

Consent, ethics and confidentiality

- f* The purpose of the proposed research is consistent with the study's own ethics approval and participant consent.
- f* There is a clear route for regulatory permissions (e.g. ethics approval) to be sought and for approvals to be notified to the study
- f* There is not an unacceptable risk to the confidentiality of the participants' identities or to other aspects of data security.
- f* There is a clear policy on whether, when and how re-identification of participants may be considered, and under what controls.
- f* Sharing with organisations outside the UK conforms to the DPA eighth principle that personal data are not transferred to another country outside the European Economic Area that does not have adequate protection for individuals' personal information.

Partnerships and commercial use

- f* The relationship between new user and the study team is equitable in relation to the expected benefits and management of risks.
- f* The proposed use of the data does not pose an unmanageable risk to continued participation of the cohort members.
- f* Proposed use by commercial organisations or by individuals and organisations outside the UK is for *bona-fide* research, educational or other uses is consistent with the public interest, and is on a non-exclusive basis.
- f* Commitment to publishing the findings in peer-reviewed publications is, for the purposes of this guidance, an important test of [bona fide](#) research.

Resources

- f* The study team has the expertise and capacity to support the request (including essential, ongoing facilitation).
- f* Funding is available in principle, from the study or the new user, to support the production of data for the user (or of running an analysis on their behalf).

Other criteria

- f* A study may specify other requirements, e.g. in relation to the management of intellectual property, deposition of derived data back with study, further use of derived data, and communication about the study. These will normally be a reflection of the terms of the study's standard data-sharing agreement.

Models for independent oversight of research data access

The study director/PI is responsible for the integrity of the process and for the decisions on data access. The access governance arrangements must be effective and proportionate for particular study. They must have an independent element and the key decisions must be transparent, equitable and documented. "Independent" for these purposes means 'being impartial towards decisions about the study's data' and so will exclude, amongst others, individuals who collaborate substantively with the study.

Two following models are illustrative and can be adapted by Directors/PIs to the circumstances of their own studies.

Model 1 is appropriate for studies that have or anticipate a significant sharing activity.

Model 2 is more appropriate for a study that anticipates only occasional requests, which are unlikely to give rise to scientific, technical, ethical or legal issues.

Model 1: All requests are considered by an Access Committee

- f* The committee considers all formal requests for access within an appropriate timescale, including new uses by the study team itself (beyond the funder-approved programme). It may also be tasked with reviewing the study's access policy and procedures, and with annual review of the performance in terms of outcomes and service.
- f* The committee comprises a balanced set of scientists with expertise of the purposes for which data are likely to be requested, supplemented *ad hoc* by further experts as required.
- f* There may also be roles for members with other kinds of expertise, such as legal and ethical, or experience as a study participant.
- f* The committee chairman is independent of the study (neither a co-investigator nor a collaborator). One or more other members is also independent.
- f* Experts who understand the study thoroughly can make a significant contribution to access decisions. Consequently, it may be justifiable to include members who have an interest in the study (through collaboration or otherwise using the data). Potential and actual conflicts of interest must be managed judiciously according to good practice.
- f* The committee is advisory to the study director/PI, who is not a member but attends. Other study team members are also observers and not members of the committee.

- f* The policy may provide for specified classes of “straight forward” request, considered highly likely to fulfil the criteria, to be triaged out and reviewed by a “fast/light” process and reported *post-hoc* to the committee. The policy may provide for those reviews to be conducted *ad hoc* by (i) a subset of the committee, or (ii) the study director / PI. To ensure transparency, high standards of triage, feedback and reporting to the independent committee must be maintained.

f

Model 2: Access decisions are periodically reviewed by an independent Access Advisor

- f* The study team formally reviews access requests for proposals. All significant decisions (approval, referral back for further information, and decline) are documented for subsequent independent review.
- f* An advisor (or committee) with appropriate expertise, independent of the study, is appointed to periodically review the outcomes of access requests *post hoc*. The reviewer may also be tasked to advise on a study’s access policy and procedures.
- f* Individual requests may be referred to the advisor for advice if difficult issues arise, e.g. a risk to the data, participants or study, or to depletable resources.
- f* It may be in the best interests of the study and requestors that an intention to decline a request is first referred, with a justification, to the advisor for advice.
- f* The advisor may advise the Director/PI that they need external expert advice in order to formulate their own advice to the study.
- f* The study director/PI is responsible for access decisions.

15. Questions

Who should read the policy and guidance?

The guidance has been written principally for:

- **Directors and principal investigators (PIs)** of research studies funded wholly or partly by the Medical Research Council and involving significant population or patient data collection; i.e. study leaders responsible for defining and implementing study policies and managing significant data resources
- **Informaticians, data managers, statisticians** and others responsible for the detailed processes of creating, managing, analysing and otherwise adding value to significant research data
- **Peer reviewers**, in particular MRC research Board Members, who will assess the summary Data Management Plans as well the science and make funding recommendations for research and data management.

In addition it may be helpful for:

- **Study participants** interested in how MRC aims both to promote research and respect the interests of participants in research, who generously provide information (and, often, also samples of tissue or DNA) for research for the public good.

What outcomes does MRC expect to see?

This new guidance and related activities should lead to the following outcomes:

- Greater clarity about the **resources** needed to support data-sharing, with allocation of appropriate levels of MRC funding for these activities.
- Greater **transparency** with easier discovery of valuable data collections and of the governance of access, leading to **easier access** for new research uses.
- Better **reporting** of data-sharing, with better evidence of the extent and diversity of sharing activities - and of new science through data-sharing.
- Better **recognition** of altruism by researchers and of technical innovation and other achievements in data-sharing.
- **Harmonisation** with the requirements of other funders.
- Alongside related work by MRC and its partners in public engagement, **strengthened public support** for responsible sharing and management of valuable collections of research data and tissues.
- Overall, **more and better research with greater impact**.

What force does the guidance have?

The requirement to comply with MRC's data-sharing policy forms part of the terms and conditions of an MRC grant or fellowship award, and of MRC Institute and Unit funding. In principle, they also apply to other forms of MRC funding.

Directors, principal investigators (PIs) and data managers should discuss with their MRC research Programme Manager any significant challenges in meeting the requirements so as to identify a reasonable way forward.

When is the effective start date for this policy?

Publication and review

The effective publication date of this documentation is 23 November 2011. A period of six months (until 30 June 2012) will be available for written comment. The guidance will be reviewed by MRC, taking account of comments, and the implementation experience of the research Boards, principal investigators and data managers. MRC will if necessary clarify the guidance through an amendment later during 2012.

Compliance

MRC population and patient studies in receipt of MRC funding, and that are actively collecting and/or analysing data or otherwise managing datasets with significant potential for new uses, should have taken reasonable steps to **comply with the requirements of this guidance by 31 May 2012** or **to have plans in place and action under way** to ensure compliance within a reasonable period.

Active studies do not need to have fully cleaned, coded or documented their data in order to comply with the guidance. Nevertheless, Directors/PIs are expected to plan to maximise value from the data – including through sharing - over an appropriate time period consistent with the study aims and design (i.e. with the study lifecycle). These

plans can be discussed with the MRC science programme managers and, potentially, form part of funding proposals.

For datasets that are no longer or rarely used, Directors of MRC Institutes and Units, or programmes should use their judgement as to whether and how to make the data available for sharing. The quality of research data and the metadata, their likely value for new research, and the nature of any consent of such “legacy” datasets will have a bearing on whether they are shareable and merit the necessary investment.

Relationship to summary Data Management Plans

For funding proposals submitted after 1 December 2011, the summary Data Management Plan should reflect the key elements of a study’s policy on data-sharing. Revised MRC guidance on summary DMPs will be issued in December 2011.

Directors of MRC Units and Institutes with custodianship of population and patient datasets should discuss how best to report their data-sharing performance and plans for the next quinquennium with their research MRC Programme Manager before writing their quinquennial review (QQR) report. The approach will reflect the extent and complexity of the Unit’s cohort collection.

How was this guidance developed and approved?

The overarching [policy on data-sharing](#) was approved by the MRC Council in March 2005. An MRC workshop in October 2010 on *Big Data Strategy* identified several priorities for strengthening MRC’s use of research data. They included better guidance for researchers on how to interpret MRC data-sharing policy, in particular on the governance of access. This new guidance attempts to fill that gap. MRC is currently working on the other priorities identified at the workshop, such as increasing the visibility (discoverability) of MRC data resources and improving the design and review of data management plans.

The principles of this guidance were approved by the MRC Population Health Sciences Group and commended by the PHSG to MRC’s Strategy Board in March 2011.

Recommendations on the content of the guidance were prepared for MRC Head Office under the auspices of the Data Support Service project by a consortium comprising partners from the Science and Technology Facilities Council, University College London and Oxford University. Leading MRC scientists, informaticians and data managers were consulted at several stages. MRC Head Office reworked the consortium’s final recommendations into its current form.

How do the policy and guidance relate to those of other funders?

To help harmonize research data policies, the MRC consulted an *ad hoc* group of funders that have similar, well-developed research data policies: The Wellcome Trust, the Economic and Social Research Council (ESRC), Cancer Research UK (CRUK) and the National Cancer Research Institute (NCRI). The MRC has attempted to align this guidance with that of the other funders, and with the [NCRI Template for Access Policy Development](#).

CRUK, ESRC, MRC and The Wellcome Trust are content that their data policies are aligned on all the main principles. There are some differences in detail, recognising the diversity in research community practices and different kinds of study. A particular example, the [ESRC data policy](#) requires ESRC-funded investigators to make research data available to other researchers within three months of the end of a grant. MRC by

contrast puts the onus on investigators to propose a specific policy for any [exclusive use of data](#), for peer reviewers to consider carefully as part of the funding proposal.

Approvals and funding processes to support sound data management and sharing may also differ between funders. Where studies are co-funded, principal investigators should establish which one funder is taking the lead (e.g. in peer review) and follow their guidance or consult with the respective funders. Funders reserve the right to request compliance with their policies, e.g. for cross council-funded research.

If a principal investigator finds that the policies of a stakeholder in MRC-funded research are at odds with those of the MRC (e.g. the other stakeholder requires immediate destruction of data at the end of a study), they should bring the discrepancy to the attention of their MRC Research Programme Manager.

Other related MRC policies and activities

New MRC guidance is also being developed on:

Research Data Management Plans

Updated guidance on what should be in data management plans (DMPs), and how – as part of funding proposals to MRC - they should be assessed in peer review (planned for end 2011).

Human Tissue Sharing

MRC and other UK funders will launch a shared vision for human tissue sharing, which will then be supplemented by new policy and guidance following detailed consultation with patients, the general public and research professionals. The tissues consultation has the potential to further inform the review of this guidance on research data.

The [MRC Data and Tissues Toolkit](#) provides practical guidance on compliance with ethical, statutory and other regulatory requirements for use of data (and tissue) arising from studies involving human participants.

MRC Research Data Gateway

A prototype gateway and **Population and Patient Research Data Directory** has been developed to enable the discovery of MRC-funded studies and their variables.

Open Access Publishing

MRC champions [Open Access Publishing](#) in science through its policy; the focus is primarily on published research outputs.

Freedom of Information

MRC makes information widely available primarily through publications, the website and other communications. We are committed to openness, good record-keeping and effective communication in our handling of requests for information under the [Freedom of Information Act \(2000\)](#). All information will be made freely available unless there is a good reason not to do so, in line with relevant exemptions under the Act.

Genomic and structural data

The principles in this policy guidance are consistent with the [Fort Lauderdale Principles](#) and the [Toronto Statement](#), which reflect good practice on prepublication data-sharing in large scale genomic data projects.